Exact Thresholds for Ising-Gibbs Samplers on General Graphs

Elchanan Mossel[*] Allan Sly[†]

January 27, 2010

Abstract

We establish tight results for rapid mixing of Gibbs Samplers for the Ferromagnetic Ising model on general graphs. We show that if

$$(d-1)\tanh\beta < 1,$$

then there exists a constant C such that the discrete time mixing time of Gibbs Samplers for the Ferromagnetic Ising model on *any* graph of n vertices and maximal degree d, where all interactions are bounded by β, and arbitrary external fields is bounded by $Cn\log n$. Moreover, the spectral gap is uniformly bounded away from 0 for all such graphs as well as for infinite graphs of maximal degree d.

We further show the when $d\tanh\beta < 1$, with high probability over the Erdős-Rényi random graph $G(n,d/n)$, it holds that the mixing time of Gibbs Samplers is

$$n^{1+\Theta(\frac{1}{\log\log n})}.$$

Both results are tight as it is known that the mixing time for random regular and Erdős-Rényi random graphs is, with high probability, exponential in n when $(d-1)\tanh\beta > 1$, and $d\tanh\beta > 1$, respectively. To our knowledge our results give the first tight sufficient conditions for rapid mixing of spin systems on general graphs. Moreover, our results are the first rigorous results establishing exact thresholds for dynamics on random graphs in terms of spatial thresholds on trees.

[*]Weizmann Institute and U.C. Berkeley. Supported by an Alfred Sloan fellowship in Mathematics, by NSF CAREER grant DMS-0548249 (CAREER), by DOD ONR grant (N0014-07-1-05-06), by BSF grant 2004105 and by ISF grant 1300/08

[†]Microsoft Research. Supported by NSF CAREER grant DMS-0548249 (CAREER), by DOD ONR grant (N0014-07-1-05-06) 1300/08

1 Introduction

Gibbs Sampling is a standard model in statistical physics for the temporal evolution of spin systems as well as a popular technique for sampling high dimensional distributions. The study of the convergence rate of Gibbs Samplers has thus attracted attention much attention from both statistical physics and theoretical computer science. Traditionally such systems where studied on lattices. However the applications in computer science coupled with the interest in diluted spin-glasses in theoretical physics led to an extensive exploration of properties of Gibbs sampling on general graphs of bounded degrees.

Below we will recall various definitions for measuring the convergence rate of the dynamics in spectral and total variation forms. In particular, we will used the notion of *rapid mixing* to indicate convergence in polynomial time in the size of the underlying graph.

A feature of most sufficient conditions for rapid convergence is that they either apply to general graphs, but are not (known to be) tight, or the results are known to be tight but apply only to special families of graphs, like 2-dimensional grids, or trees. Examples of results of the first type include the Dobrushin and the Dobrushin Shlosman conditions [6] and results by Vigoda and collaborators on colorings, see e.g. [28, 29, 9]. Examples of tight results for special graphs include the Ising model on 2-dimensional grids by Martinelli and Oliveri [18, 19], see also [16] and the Ising model on trees [12, 2, 20, 17].

In this paper we consider Gibbs sampling for the Ferromagnetic Ising model on general graphs and provide a criteria in terms of the maximal coupling constant β and the maximal degree d which guarantees rapid convergence for *any* graph and *any* external fields. The criteria is $(d-1)\tanh\beta < 1$. We further establish that if $d\tanh\beta < 1$, then rapid mixing holds, with high probability, on the Erdős-Rényi random graph of average degree d, thus proving the main conjecture of [23, 24]. Both results are tight as random d-regular graphs and Erdős-Rényi random graph of average degree d with no external fields, have, with high probability, mixing times that are exponential in the size of the graph when $(d-1)\tanh\beta > 1$ (resp. $d\tanh\beta > 1$) [8, 5]. To our knowledge our results are the first tight sufficient conditions for rapid mixing of spin systems on general graphs.

Our results are intimately related to the spatial mixing properties of the Gibbs measure, particularly on trees. A model has the *uniqueness property* (roughly speaking) if the marginal spin at a vertex is not affected by conditioning the spins of sets of distant vertices as the distance goes to infinity. On the infinite d-regular tree, uniqueness of the Ferromagnetic Ising model holds when $(d-1)\tanh\beta \leq 1$ [15], corresponding to the region of rapid mixing. It is known from the work of Weitz [30] that in fact spatial mixing occurs when $(d-1)\tanh\beta \leq 1$ on any graph of maximum degree d.

It is widely believed that (some form of) spatial mixing implies fast mixing of the Gibbs sampler. However, this is only known for amenable graphs and for a strong form of spatial mixing called "strong spatial mixing" [7]. While lattices are amenable, there are many ensembles of graphs which are non-amenable such as expander graphs. In fact since most graphs of bounded degree are expanders, the strong spatial mixing technique does not apply to them. Our results apply to completely general graphs and in particular various families of random graphs whose neighbourhoods have exponential growth.

Our results also immediately give lower bounds on the spectral gap of the continuous time Glauber dynamics which are independent of the size of the graph. This in turn allows us to establish lower bound on the spectral gap for the Glauber dynamics on infinite graphs of maximal degree bounded by d as well.

To understand our result related to the Erdős-Rényi random graph, we note that the threshold for the Erdős-Rényi random graphs also corresponds to a spatial mixing threshold. For a randomly chosen vertex the local graph neighbourhood is asymptotically distributed as a Galton-Watson branching process with offspring distribution Poisson with mean d. Results of Lyons [15] imply that the uniqueness threshold on the Galton-Watson tree is $d\tanh\beta < 1$ which is equal to the threshold for rapid mixing established here.

The correspondence between spatial and temporal mixing is believed to hold for many other important models. We conjecture that when there is uniqueness on the d-regular tree for the *antiferromagnetic Ising* model or the *hardcore model* then there is rapid mixing of the Gibbs sampler on all graphs of maximum degree d in these models. It is known that for both these models that the mixing time on almost all random d-regular bipartite graphs is exponential in n the size of the graph beyond the uniqueness threshold [25, 8, 5], so our conjecture is that uniqueness on the tree exactly corresponds to rapid mixing of the Gibbs sampler. We summarize our main contributions as follows:

- Our results are the first results providing tight criteria for rapid mixing of Gibbs samplers on general graphs.
- Our results show that the threshold is given by a corresponding threshold for a tree model, in particular in the case of random graphs and dilute mean field models. We note that in the theory of spin-glasses it is conjectured that for many spin systems on random diluted (bounded average degree) graphs the "dynamical threshold" for rapid mixing, is given by a corresponding "replica" threshold, i.e., a spatial threshold for a corresponding spin system on trees (see for example [21, 13, 22]). To the best of our knowledge our results are the first to rigorously establish such thresholds.

While the proof we present here is short and elegant - it is fundamentally different than previous approaches in the area. In particular:

- It is known that imitating the block dynamics technique [18, 19] cannot be extended to the non-amenable setting since the bounds rely crucially on the small boundary to volume ratio which can no be extended to expander graphs, see a more detailed discussion in [7].
- Weitz [30] noted that the tree of self avoiding walks construction establishes mixing results on amenable graphs but not for non-amenable graphs. In general, correlation inequalities/spatial mixing have previously only been shown to to imply rapid mixing on amenable graphs, an excellent reference of this is the thesis of Weitz [31].
- The technique of censoring the dynamics is another recent development in the analysis of Gibbs samplers [30] and can for instance be used to translate results on the block dynamics to those on the single site dynamics. Its standard application does not, however, yield new results for non-amenable graphs.
- While tight results have been established in the case of trees [12, 2, 20, 17] which are non-amenable, the methods do not generalize to more general graphs as they make fundamental use of properties of the tree, in particular the presence of leaves at the base. Indeed, the fact that the median degree of a tree is 1 illustrates the difference between trees and regular graphs.

The main novelty in our approach is a new application of the censoring technique. In the standard use of censoring a censored Markov chain is constructed which is shown to mix rapidly and then the censoring inequality implies rapid mixing of the original dynamics. Our approach is a subtle conceptual shift. Rather than construct a censoring scheme which converges to the stationary distribution we construct a sequence of censored dynamics which do not converge to stationarity. They do, however, allow us to establish a sequence of recursive bounds from which we derive our estimates of the spectral gap and the mixing time.

Another serious technical challenge of the paper was determining the correct mixing time for the Gibbs sampler on Erdős-Rényi random graphs. The necessary estimate is to bound the mixing time on the local neighbourhoods of the graph which are Galton-Watson branching processes with Poisson offspring distribution. This is done via an involved distributional recursive analysis of the cutwidth of these branching process trees.

In the following subsections we state our results, then recall the definition of the Ising model, Gibbs Sampling and Erdős-Rényi random graphs followed by a statement of a general theorem from which both of our

main results follow. We then sketch the main steps of the proof followed by detailed proofs. We then show how our spectral gap bounds on finite graphs can be extended to infinite graphs. Finally we conclude with open problems involving other systems.

1.1 Our Results

In our main result we establish the following tight criteria for rapid mixing of Gibbs sampling for general graphs in terms of the maximal degree.

Theorem 1 *For any integer $d \geq 2$, and inverse temperature $\beta > 0$, such that*

$$(d-1)\tanh\beta < 1, \tag{1}$$

there exist constants $0 < \lambda^*(C,\beta), C(d,\beta) < \infty$, such that on any graph of maximum degree d on n vertices, the discrete time mixing time of the Gibbs sampler for the ferromagnetic Ising model with all edge interactions bounded by β, and arbitrary external fields, is bounded above by $Cn\log n$.

Further the continuous time spectral gap of the dynamics is bounded below by λ^*. The spectral gap bound applies also for infinite graphs.

We note that a lower bound of $\Omega(n\log n)$ on the mixing time follows from the general results of [10].

The techniques we develop here also allow us to derive results for graphs with unbounded degrees. Of particular interest is the following tight result:

Theorem 2 *Let $\beta > 0$ and $d > 0$ and consider the Erdős-Rényi random graph G on n vertices where each edge is present independently with probability d/n. Then for all β such that $d\tanh\beta < 1$, there exists $c(d,\beta)$ and $C(d,\beta)$, such that with high probability over G, the discrete time mixing time τ_{mix} of the Gibbs sampler for the ferromagnetic Ising model with all edge interactions bounded by β and arbitrary external field satisfies*

$$n^{\left(1+\frac{c}{\log\log n}\right)} \leq \tau_{mix} \leq n^{\left(1+\frac{C}{\log\log n}\right)}$$

while the continuous time while spectral gap satisfies

$$n^{-\frac{c}{\log\log n}} \geq \mathit{Gap} \geq n^{-\frac{C}{\log\log n}}.$$

Both results are tight as estimates obtained in [8, 5] following [25] and proving a conjecture from [23, 24] imply that for the Ising model without external fields, the mixing time of the Gibbs sampler is with high probability $\exp(\Omega(n))$ on random d-regular graphs if $(d-1)\tanh\beta > 1$ and Erdős-Rényi random graphs of average degree d when $d\tanh\beta > 1$.

1.2 Standard Background

In the following subsection we recall some standard background on the Ising Model, Gibbs Sampling and Erdős-Rényi Random Graphs.

1.2.1 The Ising Model

The Ising model is perhaps the oldest and simplest discrete spin system defined on graphs. This model defines a distribution on labelings of the vertices of the graph by $+$ and $-$.

Definition 1 *The (homogeneous) Ising model on a graph G with inverse temperature β is a distribution on configurations $\{\pm\}^V$ such that*

$$P(\sigma) = \frac{1}{Z(\beta)} \exp(\beta \sum_{\{v,u\} \in E} \sigma(v)\sigma(u)) \tag{2}$$

where $Z(\beta)$ is a normalizing constant.

More generally, we will be interested in the more general Ising models defined by:

$$P(\sigma) = \frac{1}{Z(\beta)} \exp(H(\sigma)), \tag{3}$$

where the Hamiltonian $H(\sigma)$ is defined as

$$H(\sigma) = \sum_{\{v,u\} \in E} \beta_{u,v} \sigma(v)\sigma(u) + \sum_v h_v \sigma(v)$$

and where h_v are arbitrary and $\beta_{u,v} \geq 0$ for all u and v. In the more general case we will write $\beta = \max_{u,v} \beta_{u,v}$.

1.2.2 Gibbs Sampling

The Gibbs sampler (also Glauber dynamics or heat bath) is a Markov chain on configurations where a configuration σ is updated by choosing a vertex v uniformly at random and assigning it a spin according to the Gibbs distribution conditional on the spins on $G - \{v\}$.

Definition 2 *Given a graph $G = (V, E)$ and an inverse temperature β, the Gibbs sampler is the discrete time Markov chain on $\{\pm\}^V$ where given the current configuration σ the next configuration σ' is obtained by choosing a vertex v in V uniformly at random and*

- *Letting $\sigma'(w) = \sigma(w)$ for all $w \neq v$.*
- *$\sigma'(v)$ is assigned the spin $+$ with probability*

$$\frac{\exp(h_v + \sum_{u:(v,u) \in E} \beta_{u,v} \sigma(u))}{\exp(h_v + \sum_{u:(v,u) \in E} \beta_{u,v} \sigma(u)) + \exp(-h_v - \sum_{u:(v,u) \in E} \beta_{u,v} \sigma(u))}.$$

We will be interested in the time it takes the dynamics to get close to the distributions (2) and (3). The *mixing time* τ_{mix} of the chain is defined as the number of steps needed in order to guarantee that the chain, starting from an arbitrary state, is within total variation distance $1/2e$ from the stationary distribution. The mixing time has property that for any integer k and initial configuration x,

$$\|P(X_{k\tau_{mix}} = \cdot \mid X_0 = x) - P(\cdot)\|_{\mathrm{TV}} \leq e^{-k}. \tag{4}$$

It is well known that Gibbs sampling is a reversible Markov chain with stationary distribution P. Let $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_m \geq -1$ denote the eigenvalues of the transition matrix of Gibbs sampling. The *spectral gap* is denoted by $\min\{1 - \lambda_2, 1 - |\lambda_m|\}$ and the *relaxation time* τ is the inverse of the spectral gap. The relaxation time can be given in terms of the Dirichlet form of the Markov chain by the equation

$$\tau = \sup \left\{ \frac{2\sum_\sigma P(\sigma)(f(\sigma))^2}{\sum_{\sigma \neq \tau} Q(\sigma, \tau)(f(\sigma) - f(\tau))^2} : \sum_\sigma P(\sigma) f(\sigma) \neq 0 \right\} \tag{5}$$

where $f : \{\pm\}^V \to \mathbb{R}$ is any function on configurations, $Q(\sigma, \tau) = P(\sigma)P(\sigma \to \tau)$ and $P(\sigma \to \tau)$ is transition probability from σ to τ. We use the result that for reversible Markov chains the relaxation time satisfies

$$\tau \leq \tau_{mix} \leq \tau \left(1 + \frac{1}{2}\log(\min_{\sigma} P(\sigma)^{-1})\right) \tag{6}$$

where τ_{mix} is the mixing time (see e.g. [1]) and so by bounding the relaxation time we can bound the mixing time up to a polynomial factor.

While our results are given for the discrete time Gibbs Sampler described above, it will at times be convenient to consider the continuous time version of the model. Here sites are updated at rate 1 by independent Poisson clocks. The two chains are closely related, the relaxation time of the continuous time Markov chain is n times the relaxation time of the discrete chain (see e.g. [1]).

1.2.3 Erdős-Rényi Random Graphs and Other Models of graphs

The Erdős-Rényi random graph $G(n, p)$, is the graph with n vertices V and random edges E where each potential edge $(u, v) \in V \times V$ is chosen independently with probability p. We take $p = d/n$ where $d \geq 1$ is fixed. In the case $d < 1$, it is well known that with high probability all components of $G(n, p)$ are of logarithmic size which implies immediately that the dynamics mix in polynomial time for all β. A random d-regular graph $\mathcal{G}(n, d)$ is a graph uniformly chosen from all d-regular graphs on n labeled vertices.

Asymptotically the local neighbourhoods of $G(n, d/n)$ and $\mathcal{G}(n, d)$ are trees. In the later case it is a tree where every node has exactly $d-1$ offsprings (except for the root which has d off-springs). In the former case it is essentially a Galton-Watson branching process with offspring distribution which is essentially Poisson with mean $d - 1$. Recall that the tree associated with a Galton-Watson branching process with offspring distribution X is a random rooted tree defined as follows: for every vertex in the tree its number of offspring vertices is independent with distribution X.

1.3 A General Theorem

Theorems 1 and 2 are both proved as special cases of the following theorem which may be of independent interest. For a graph $G = (V, E)$ and vertex $v \in V$, we write $B(v, R)$ for the ball of radius R around v, i.e., the set of all vertices that are of distance at most R from v. We write $S(v, R) = B(v, R) \setminus B(v, R - 1)$ for the sphere of radius R around v.

Theorem 3 *Let G be a graph on $n \geq 2$ vertices such that there exist constants $R, T, \mathfrak{X} \geq 1$ such that the following three conditions holds for all $v \in V$:*

- ***Volume:*** *The volume of the ball $B(v, R)$ satisfies $|B(v, R)| \leq \mathfrak{X}$.*
- ***Local Mixing:*** *For any configuration η on $S(v, R)$ the continuous time mixing time of the Gibbs sampler on $B(v, R - 1)$ with fixed boundary condition η is bounded above by T.*
- ***Spatial Mixing:*** *For each vertex $u \in S(v, R)$ define*

$$a_u = \sup_{\eta^+, \eta^-} P(\sigma_v = +|\sigma_\Lambda = \eta^+) - P(\sigma_v = +|\sigma_\Lambda = \eta^-) \tag{7}$$

where the supremum is over configurations η^+, η^- on $S(v,R)$ which differ only at u with $\eta_u^+ = +, \eta_u^- = -$. Then

$$\sum_{u \in S(v,R)} a_u \le \frac{1}{4}. \tag{8}$$

Then starting from the all $+$ and all $-$ configurations in continuous time the monotone coupling couples with probability at least $\frac{7}{8}$ by time $T \lceil \log 8\mathfrak{X} \rceil (3 + \log_2 n)$.

It follows that the mixing time of the Gibbs sampler in continuous time satisfies

$$\tau_{mix} \le T \lceil \log 8\mathfrak{X} \rceil (3 + \log_2 n)$$

while the spectral gap satisfies

$$\textit{Gap} \ge (T \lceil \log 8\mathfrak{X} \rceil)^{-1} \log 2.$$

We will write $\mathrm{Vol}(R, \mathfrak{X})$ for the statement that $|B(v,R)| \le \mathfrak{X}$ for all $v \in V$, write $\mathrm{SM}(R)$ for the statement that (8) holds for all $v \in V$ and write $\mathrm{LM}(R,T)$ for the statement that the continuous time mixing time of the Gibbs sampler on $B(v, R-1)$ is bounded above by T for any fixed boundary condition η . Using this notation the theorem states that:

$$\mathrm{Vol}(R, \mathfrak{X}) \text{ and } \mathrm{SM}(R) \text{ and } \mathrm{LM}(R,T) \implies \tau_{mix} \le T \lceil \log 8\mathfrak{X} \rceil (3 + \log_2 n). \tag{9}$$

In the conclusion section of the paper we state a much more general version of Theorem 3 which applies to general monotone Gibbs distributions and allows the sets $B(v,R)$ to be arbitrary sets containing v (where $S(v,R)$ is replaced by the inner vertex boundary of the $B(v,R)$). We note that the implication proven here for monotone systems showing

$$\text{Spatial Mixing} \implies \text{Temporal Mixing}$$

is stronger than that established in previous work [27, 18, 3, 7] where it is shown that Strong Spatial Mixing implies Temporal Mixing for graphs with sub-exponential growth (Strong Spatial Mixing says that the quantity a_u decays exponentially in the distance between u and v). In particular, Theorem 3 applies also to graphs with exponential growth and for a very general choice of blocks $B(v,R)$. Both Theorems 1 and 2 deal with expanding graphs where Theorem 3 is needed.

A different way to look at our result is as a strengthening of the Dobrushin-Shlosman condition [6]. Stated in its strongest form in [31] Theorem 2.5, it says that rapid mixing occurs if the effect on the spin at a vertex v of disagreements on the boundary of blocks containing v is small - averaged over all blocks containing v - then the model has uniqueness and the block dynamics mixes rapidly. Theorem 4 requires only that for each vertex there exists a block such that the boundary effect is small. This is critical in expanders and random graphs where the boundary of a block is proportional to its volume.

1.4 Proofs Sketch

We briefly discuss the main ideas in our proofs of Theorems 3,1 and 2.

1.4.1 Theorem 3 and Censoring

The proof of Theorem 3 is based on considering the monotone coupling of the continuous time dynamics starting with all $+$ and all $-$ states and showing that there exists a constant s such that at time ks, for all vertices v, the probability that the two measures have not coupled at v is at most 2^{-k}.

In order to prove such a claim by induction, it is useful to censor the dynamics from time ks onwards by not performing any updates outside a ball of radius R around v. Recent results of Peres and Winkler show that doing so will result in a larger disagreement probability at v than without any censoring.

For the censored dynamics we use the triangle inequality and compare the marginal probability at v for the two measures by comparing each distribution to the stationary distribution at v given the boundary condition and then comparing the two stationary distributions at v given the two boundary conditions.

By using $\mathrm{LM}(R,T)$ and running the censored dynamics for $T\lceil \log 8\mathfrak{X}\rceil$ time we can ensure that the error of the first type contributes at most $2/(8\mathfrak{X})$ in case where the two boundary conditions are different and therefore at most $2/(8\mathfrak{X})$ times the expected number of disagreements at the boundary which is bounded by 2^{-k-2} by induction. By using $\mathrm{SM}(R)$ and the induction hypothesis we obtain that the expected discrepancy between the distributions at σ_v given the two different boundary conditions is at most 2^{-k-2}. Combining the two estimates yields the desired result. As this gives an exponential rate of decay in the expected discrepancy it establishes a constant lower bound on the spectral gap.

The proofs of Theorems 1 and 2 follows from (9) by establishing bounds on $\mathrm{Vol}, \mathrm{SM}$ and LM.

1.4.2 Bounding the Volume

The easiest step in both Theorems 1 and 2 is to establish $\mathrm{Vol}(R,\mathfrak{X})$. For graphs of degree at most d, the volume grows as $O((d-1)^R)$ and using arguments from [24] one can show that if $R = (\log\log n)^n$ then for $G(n,d/n)$ one can take $\mathfrak{X}$ of order $d^R \log n$.

1.4.3 Spatial Mixing Bounds

Establishing Spatial mixing bounds relies on the fact that for trees without external fields - this is a standard calculation. The presence of external fields can be dealt with using a Lemma from [2] which shows that the for Ising model on trees, the difference in magnetization is maximized when there are no external fields. A crucial tool which allows us to obtain results for non-tree graphs is the Weitz tree [30]. This tree allows us to write magnetization ratios for the Ising model on general graphs using a related model on the tree. In [24] it was shown that the Weitz tree can be used to construct an efficient algorithm different than Gibbs Sampling for sampling Ising configurations under the conditions of Theorems 1 and 2 (the running time of the algorithm is $n^{1+C(\beta)}$ compared to $C(\beta) n \log n$ established here).

1.4.4 Local Mixing Bounds

In order to derive local mixing bounds we generalize results from [2] on the mixing times in terms of cut-width to deal with arbitrary external fields. Further, for the case of Erdős-Rényi random graphs and $R = (\log\log n)^2$ we show that with high probability the cut width is of order $\log n / \log\log n$.

2 Proofs

In this section we prove Theorems 3, 1 and 2 while the verification of the $\mathrm{Vol}, \mathrm{SM}$ and LM conditions is deferred to the following sections. We begin by recalling the notion of monotone coupling and the result by Peres-Winkler on censoring. We then proceed with the proof of the theorems.

2.1 Monotone Coupling

For two configurations $X, Y \in \{-,+\}^V$ we let $X \geq Y$ denote that X is greater than or equal to Y pointwise. When all the interactions β_{ij} are positive, it is well known that the Ising model is a monotone system under this partial ordering, that is if $X \geq Y$ then,

$$P\left(\sigma_v = +|\sigma_{V\setminus\{v\}} = X_{V\setminus\{v\}}\right) \geq P\left(\sigma_v = +|\sigma_{V\setminus\{v\}} = Y_{V\setminus\{v\}}\right).$$

As it is a monotone system there exists a coupling of Markov chains $\{X_t^x\}_{x\in\{-,+\}^V}$ such that marginally each has the law of the Gibbs Sampler with starting configurations $X_0^x = x$ and further that if $x \geq y$ then for all t, $X_t^x \geq X_t^y$. This is referred to as the monotone coupling and can be constructed as follows: let $v_1, \ldots$ be a random sequence of vertices updated by the Gibbs Sampler and associate with them iid random variables $U_1, \ldots$ distributed as $U[0,1]$ which determine how the site is updated. At the ith update the site v_i is updated to $+$ if

$$U_i \leq \frac{\exp(h_v + \sum_{u:(v,u)\in E} \beta_{u,v}\sigma(u))}{\exp(h_v + \sum_{u:(v,u)\in E} \beta_{u,v}\sigma(u)) + \exp(-h_v - \sum_{u:(v,u)\in E} \beta_{u,v}\sigma(u))}$$

and to $-$ otherwise. It is well known that such transitions preserve the partial ordering which guarantees that if $x \geq y$ then $X_t^x \geq X_t^y$ by the monotonicity of the system. In particular this implies that it is enough to bounded the time taken to couple from the all $+$ and all $-$ starting configurations.

2.2 Censoring

In general it is believed that doing more updates should lead to a more mixed state. For the ferromagnetic Ising model and other monotone systems this intuition was proved by Peres and Winkler. They showed that starting from the all $+$ (or all $-$) configurations adding updates only improves mixing. More formally they proved the following proposition.

Proposition 1 *Let $u_1, \ldots, u_m$ be a sequence of vertices and let $i_1, \ldots, i_l$ be a strictly increasing subsequence of $1, \ldots, m$. Let X^+ (resp. X^-) be a random configuration constructed by starting from the all $+$ (resp. all $-$) configuration and running Gibbs updates sequentially on $u_1, \ldots, u_m$. Similarly let Y^+ (resp. Y^-) be a random configuration constructed by starting from the all $+$ (resp. all $-$) configuration and running Gibbs updates sequentially on the vertices $u_{i_1}, \ldots, u_{i_m}$. Then*

$$Y^- \preccurlyeq X^- \preccurlyeq X^+ \preccurlyeq Y^+.$$

where $A \preccurlyeq B$ denotes that A stochastically dominates B in the partial ordering of configurations.

This result in fact holds for random sequences of vertices of random length and random subsequences provided the choice of sequence is independent of the choices that the Gibbs sampler makes. The result remains unpublished but its proof can be found in [26].

2.3 Proof of Theorem 3

Proof:[***Theorem 3***]

Let X_t^+, X_t^-, denote the Gibbs sampler on G started respectively from the all $+$ and $-$ configurations, coupled using the monotone coupling described in Section 2.1. Fix some vertex $v \in G$. We will define two new censored chains Z_t^+ and Z_t^- starting from the all $+$ and all $-$ configurations respectively. Take

$S \geq 0$ to be some arbitrary constant. Until time S we set both Z_t^+ and Z_t^- to be simply equal to X_t^+ and X_t^- respectively. After time S all updates outside of $B(v, R-1)$ are censored, that is Z_t^+ and Z_t^- remain unchanged on $V \setminus B(v, R-1)$ after time S but inside $B(v, R-1)$ share all the same updates with X_t^+ and X_t^-.

In particular this means that for Z_t^+ and Z_t^- the spins on $S(v, R)$ are fixed after time S. By monotonicity of the updates we have $Z_t^+ \geq Z_t^-$ and $X_t^+ \geq X_t^-$ for all t. After time S the censored processes are simply the Gibbs sampler on $B(v, R-1)$ with boundary condition $X_S^\pm(S(v, R))$. By assumption we have that the mixing time of this dynamics is bounded above by T and by equation (4) if $t = T\lceil \log 8\mathfrak{X} \rceil$ then

$$\left| P\left(Z_{S+t}^+(v) = + \mid \mathcal{F}_S\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^+(S(v, R))\right) \right| \leq \frac{1}{8\mathfrak{X}}. \tag{10}$$

and similarly for Z^- where $\mathcal{F}_S$ denotes the sigma-algebra generated by the updates up to time S. Now

$$\begin{aligned} &P\left(Z_{S+t}^+(v) \neq Z_{S+t}^-(v) \mid \mathcal{F}_S\right) = P\left(Z_{S+t}^+(v) = + \mid \mathcal{F}_S\right) - P\left(Z_{S+t}^-(v) = + \mid \mathcal{F}_S\right) \\ &\quad = I\left(X_S^+(B(v, R)) \neq X_S^-(B(v, R))\right) \left[P\left(Z_{S+t}^+(v) = + \mid \mathcal{F}_S\right) - P\left(Z_{S+t}^-(v) = + \mid \mathcal{F}_S\right)\right], \end{aligned} \tag{11}$$

since if $X_S^+(B(v, R)) = X_S^-(B(v, R))$ then the censored processes remain equal within $B(v, R)$ for all time as they receive the same updates. Now we split up the right hand side as follows so by the triangle inequality

$$\begin{aligned} &I\left(X_S^+(B(v, R)) \neq X_S^-(B(v, R))\right) \left[P\left(Z_{S+t}^+(v) = + \mid \mathcal{F}_S\right) - P\left(Z_{S+t}^-(v) = + \mid \mathcal{F}_S\right)\right] \\ &\leq I\left(X_S^+(B(v, R)) \neq X_S^-(B(v, R))\right) \Bigg[\left| P\left(Z_{S+t}^+(v) = + \mid \mathcal{F}_S\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^+(S(v, R))\right) \right| \\ &\qquad + \left| P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^+(S(v, R))\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^-(S(v, R))\right) \right| \\ &\qquad + \left| P\left(Z_{S+t}^-(v) = + \mid \mathcal{F}_S\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^-(S(v, R))\right) \right| \Bigg]. \end{aligned} \tag{12}$$

Now

$$\begin{aligned} &EI\left(X_S^+(B(v, R)) \neq X_S^-(B(v, R))\right) \left| P\left(Z_{S+t}^+(v) = + \mid \mathcal{F}_S\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^+(S(v, R))\right) \right| \\ &\quad \leq \frac{1}{8\mathfrak{X}} EI\left(X_S^+(B(v, R)) \neq X_S^-(B(v, R))\right) \\ &\quad \leq \frac{1}{8\mathfrak{X}} \sum_{u \in B(v,R)} P\left(X_S^+(u) \neq X_S^-(u)\right) \\ &\quad \leq \frac{1}{8} \max_{u \in V} P\left(X_S^+(u) \neq X_S^-(u)\right) \end{aligned} \tag{13}$$

where the second inequality follows from equation (10) and the final inequality follows from the volume assumption. Similarly for Z^-.

If $\eta^+ \geq \eta^-$ are two configurations on $S(v, R)$ which differ only on the set $U \subseteq S(v, R)$ then by changing the vertices one at a time by the spatial mixing condition we have that

$$P(\sigma_v = + | \sigma_\Lambda = \eta^+) - P(\sigma_v = + | \sigma_\Lambda = \eta^-) \leq \sum_{u \in U} a_u$$

It follows that

$$\begin{aligned} &E\left| P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^+(S(v, R))\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} = X_S^-(S(v, R))\right) \right| \\ &\quad \leq \mathrm{E} \sum_{u \in B(v,R)} a_u I\left(X_S^+(u) \neq X_S^-(u)\right) \leq \frac{1}{4} \max_{u \in V} P\left(X_S^+(u) \neq X_S^-(u)\right). \end{aligned} \tag{14}$$

Combining equations (11), (12), (13) and (14) we have that

$$P\left(Z^+_{S+t}(v) \neq Z^-_{S+t}(v)\right) \leq \frac{1}{2}\max_{u\in V} P\left(X^+_S(u) \neq X^-_S(u)\right).$$

By the Censoring Lemma we have that $Z^+_t \succcurlyeq X^+_t \succcurlyeq X^-_t \succcurlyeq Z^-_t$ and so,

$$P\left(X^+_{S+t}(v) \neq X^-_{S+t}(v)\right) \leq P\left(Z^+_{S+t}(v) \neq Z^-_{S+t}(v)\right)$$

Combining the previous two equations and taking a maximum over v we have that

$$\max_{u\in V} P\left(X^+_{S+t}(u) \neq X^-_{S+t}(u)\right) \leq \frac{1}{2}\max_{u\in V} P\left(X^+_S(u) \neq X^-_S(u)\right). \tag{15}$$

Now S is arbitrary so iterate equation (15) to get that

$$\max_{u\in V} P\left(X^+_{t(3+\lceil\log_2 n\rceil)}(u) \neq X^-_{t(3+\lceil\log_2 n\rceil)}(u)\right) \leq 2^{-3-\lceil\log_2 n\rceil} \leq \frac{1}{2en}.$$

Taking a union bound over all $u \in V$ we have that

$$P\left(X^+_{t(3+\lceil\log_2 n\rceil)} \not\equiv X^-_{t(3+\lceil\log_2 n\rceil)}\right) \leq \frac{1}{2e}$$

and so the mixing time is bounded above by $T\lceil\log 8\mathfrak{X}\rceil\,(3+\log_2 n)$. Since the expected number of disagreements decays exponentially with a rate of at least $t^{-1}\log 2$, i.e.

$$E\#\left\{u \in V : X^+_s(u) \neq X^-_s(u)\right\} \leq 2ne^{-st^{-1}\log 2}$$

it follows by standard results (see e.g. [4]) that the spectral gap is bounded below by $t^{-1}\log 2$. ■

2.4 Proofs of Theorems 1 and 2

We now prove Theorems Theorems 1 and 2 except for the result for infinite graphs which will be proven in Section 6. Theorem 1 follows from (9) and the following lemmas.

Lemma 1 *Let $G = (V, E)$ be a graph of maximal degree d. Then $\mathrm{Vol}(R, \mathfrak{X})$ holds with*

$$\mathfrak{X} = 1 + d\sum_{\ell=1}^{R}(d-1)^{\ell-1}.$$

Lemma 2 *Let $G = (V, E)$ be a graph of maximal degree d and consider the ferromagnetic Ising model on G with arbitrary external fields. Then $\mathrm{LM}(R, T)$ holds with*

$$T = 80d^3\mathfrak{X}^3 e^{5\beta d(\mathfrak{X}+1)}, \quad \mathfrak{X} = 1 + d\sum_{\ell=1}^{R}(d-1)^{\ell-1}.$$

Lemma 3 *Let $G = (V, E)$ be a graph with maximum degree d and let $v \in V$. Suppose that $(d-1)\tanh\beta < 1$. Let R be an integer large enough so that*

$$\frac{d(d-1)^{R-1}\tanh^R\beta}{1-(d-1)\tanh\beta} \leq \frac{1}{4}. \tag{16}$$

Then $\mathrm{SM}(R)$ holds.

We note that Lemma 1 is trivial. As for Lemma 2 - it is easy to prove a bound with a finite T depending on R only assuming all external fields are bounded. We provide an analysis with a tighter bound which applies also when the external fields are not bounded. The proof is based on cut-width. The main step is proving Lemma 3 which uses recursions on trees, a comparison argument from [2] and the Weitz tree.

The upper bound in Theorem 2 follows from (9) and the following lemmas.

Lemma 4 *Let G be a random graph distributed as $G(n, d/n)$. Then $\mathrm{Vol}(R, \mathfrak{X})$ holds with high probability over G with*

$$R = (\log\log n)^2, \quad \mathfrak{X} = d^R \log n.$$

Lemma 5 *Let G be a random graph distributed as $G(n, d/n)$ where d is fixed. There exists a constant $C(d)$ such that for $\mathrm{LM}(R, T)$ holds with high probability over G with*

$$R = (\log\log n)^2, \quad T = e^{10\beta C(d) \frac{\log n}{\log\log n})}.$$

Lemma 6 *Let G be a random graph distributed as $G(n, d/n)$ where d is fixed and $d \tanh \beta < 1$. Then $\mathrm{SM}(R, T)$ holds with high probability over G with $R = (\log\log n)^2$.*

The main challenge in extending the proof from bounded degree graphs to $G(n, d/n)$ is obtaining a good enough control on the local geometry of the graph. In particular, we obtain very tight tail estimates on the cut-width of a Galton-Watson tree with Poisson offspring distribution of $(\log\log n)^2$ levels. A lower bound on the mixing time of $n^{1+\Omega(\frac{1}{\log\log n})}$ was shown in [24] by analyzing large star subgraphs on $G(n, d/n)$. Recall that a star is a graph which is a rooted tree with depth 1 and that an Erdős-Rényi random graph with high probability there are stars with degree $\Omega(\frac{\log n}{\log\log n})$.

3 Volume Growth

We begin with verification of the Volume Growth condition. Since Lemma 1 is trivial, this section will be devoted to the proof of Lemma 4 and other geometric properties of random graphs. The reader who is interested in the proof of Theorem 1 only may skip the remainder of this section.

The results stated in the section will require the notion of *tree excess*. For a graph G we let $t(G)$ denote the *tree excess* of G, i.e.,

$$t(G) = |E| - |V| + 1.$$

Note that the second item of the following lemma implies the statement of Lemma 4.

Lemma 7 *Let d be fixed and let G be a random graph distributed as $G(n, d/n)$. The following hold with high probability over G when $R = (\log\log n)^2$ for all $v \in G$:*

- *$B(v, R)$ has a spanning tree $T(v, R)$ which is stochastically dominated by a Galton-Watson branching process with offspring distribution Poisson(d).*
- *The tree excess satisfies $t(v, R) \leq 1$.*
- *The volume of $B(v, R)$ is bounded by*

$$|B(v, R)| \leq d^R \log n.$$

Proof: We construct a spanning tree $T(v,R)$ of $B(v,R)$ in a standard manner. Take some arbitrary ordering of the vertices of G. Start with the vertex v and attach it to all its neighbors in G. Now take the minimal vertex in $S(v,1)$, according to the ordering, and attach it to all its neighbors in G which are not already in the tree. Repeat this for each of the vertices in $S(v,1)$ in increasing order. Repeat this for $S(v,2)$ and continue until $S(v,R-1)$ which completes $T(v,R)$. By construction this is a spanning tree for $B(v,R)$. The construction can be viewed as a breadth first search of $B(v,R)$ starting from v and exploring according to the vertex ordering. By a standard argument $T(v,R)$ is stochastically dominated by a Galton-Watson branching process with offspring distribution Poisson(d) with R levels thus proving the first statement.

Since the volume of $B(v,R)$ equals the volume of $T(v,R)$ it suffices to bound the later. For this we use a variant of an argument from [24]. We let $Z(r)$ denote the distribution of the volume of a Galton Watson tree of depth r with off spring distribution N where N is Poisson(d). We claim that for all $t > 0$ it holds that

$$\sup_r E[\exp(tZ_r d^{-r})] < \infty. \tag{17}$$

Writing $s = s(t)$ for the value of the supremum, if follows from Markov'v inequality that,

$$s \geq P[Z_R \geq R^d \log n] \exp(t \log n)$$

and so

$$P[Z_R \geq R^d \log n] \leq s \exp(-t \log n),$$

which is smaller than $o(1/n)$ if $t > 1$. This implies that $B(v,R) \leq R^d \log n$ for all v by a union bound and proves the second statement of the lemma.

For (17), let N_i be independent copies of N and note that

$$\begin{aligned} E\exp(tZ_{r+1}) &= E\exp(\sum_{i=0}^{Z_r} td^{-(r+1)}N_i) = E[E[\exp(\sum_{i=0}^{Z_r} td^{-(r+1)}N_i|Z_r]] \\ &= E[(E[\exp(td^{-r+1}N)])^{Z_r}] = E\exp(Z_r \log(E\exp(td^{-(r+1)}N))), \end{aligned} \tag{18}$$

which recursively relates the exponential moments of Z_{r+1} to the exponential moments of Z_r. In particular since all the exponential moments of Z_1 exist, $E\exp(tZ_r) < \infty$ for all t and r. When $0 < s \leq 1$

$$E\exp(sN) = \sum_{i=0}^{\infty} \frac{s^i EN^i}{i!} \leq 1 + sd + s^2 \sum_{i=2}^{\infty} \frac{EN^i}{i!} \leq \exp(sd(1+\alpha s)) \tag{19}$$

provided α is sufficiently large. Now fix a t and let $t_r = t\exp(2\alpha t \sum_{i=r+1}^{\infty} d^{-i})$. For some sufficiently large j we have that $\exp(2\alpha t \sum_{i=r+1}^{\infty} d^{-i}) < 2$ and $t_r d^{-(r+1)} < 1$ for all $r \geq j$. Then for $r \geq j$ by equations (18) and (19),

$$\begin{aligned} E\exp(t_{r+1}Z_{r+1}d^{-(r+1)}) &= E\exp(\log(E\exp(t_{r+1}d^{-(r+1)}N_i))Z_r) \\ &\leq E\exp(t_{r+1}(1+\alpha t_{r+1}d^{-(r+1)})Z_r d^{-r}) \\ &\leq E\exp(t_{r+1}(1+2\alpha t d^{-(r+1)})Z_r d^{-r}) \\ &\leq E\exp(t_r Z_r d^{-r}) \end{aligned}$$

and so

$$\sup_{r\geq j} E\exp(tZ_r d^{-r}) \leq \sup_{n\geq j} E\exp(t_r Z_r d^{-r}) = E\exp(t_j Z_j d^{-j}) < \infty$$

which completes the proof of (17).

It remains to bound the tree excess. In the construction of $T(v,R)$ there may be some edges in $B(v,R)$ which are not explored and so are not in $T(v,R)$. Each edge between $u, w \in V(v,R)$ which is not explored in

the construction of $T(v,R)$ is present in $B(v,R)$ independently with probability d/n. There are at most d^{2R} unexplored edges and

$$P(\text{Binomial}(d^{2R}, d/n) > 1] \leq d^{4R}(d/n)^2 \leq n^{-2+o(1)},$$

for any fixed d. So by a union bound with high probability we have that $t(v,R) \leq 1$ for all v. ■

4 Local Mixing

In this section we prove Lemma 2 and Lemma 5. The proof that the local mixing condition holds for graphs of bounded degree, bounded volume and bounded external field is standard. Indeed the reader who is interested in Theorem 1 for models with bounded external fields may skip this section.

4.1 Cut-Width Bounds

The main tool in bounding the mixing time will be the notion of cut-width used in [2]. Recall that the *cut-width* of a finite graph $G = (V, E)$

$$\min_{\pi \in S(n)} \max_{1 \leq i \leq n-1} |\{v_{\pi(j)} : j \leq i\} \times \{v_{\pi(j)} : j > i\} \cap E|$$

where the minimum is taken over all permutations of the labels of the vertices $v_1, \ldots, v_n$ in V.

We will prove the following result which generalizes the results of [2] to the case with boundary conditions. The proof follow the ones given in [2] and [16].

Lemma 8 *Consider the Ising model on G with interaction strengths bounded by β, arbitrary external field, cut-width $\mathcal{E}$, and maximal degree d. Then the relaxation time of the discrete time Gibbs sampler is at most $n^2 e^{4\beta(\mathcal{E}+d)}$.*

Proof: We follow the notation of [12]. Fix an ordering "$<$" of the vertices in V which achieves the cut-width. Define a canonical path $\gamma(\sigma, \eta)$ between two configurations σ, η as follows: let $v_1 < v_2 < \ldots < v_\ell$ be the vertices on which σ and η differ. The kth configuration in the path $\eta = \sigma^{(0)}, \sigma^{(1)}, \ldots, \sigma^{(\ell)}$ is defined by $\sigma^{(k)}_v = \sigma_v$ for $v \leq v_k$ and $\sigma^{(k)}_v = \eta_v$ for $v > v_k$. Then by the method of canonical paths (see e.g. [11, 16]) the relaxation time is bounded by

$$\tau \leq n \sup_e \sum_{\sigma,\eta: e \in \gamma(\sigma,\eta)} \frac{P(\sigma)P(\eta)}{Q(e)}$$

where the supremum is over all pairs of configurations $e = (x, y)$ which differ at a single vertex and where $e \in \gamma(\sigma, \eta)$ denotes that x and y are consecutive configurations in the canonical path $\gamma(\sigma, \eta)$ and $Q((x, y)) = P(x)P(x \to y)$.

Let $e = (x, y)$ be a pair of configurations which differ only at v. For a pair of configurations σ, η let $\varphi_e(\sigma, \eta)$ denote the configuration which is given by $\varphi_e(\sigma, \eta)_{v'} = \eta_{v'}$ for $v' < v$ and $\varphi_e(\sigma, \eta)_{v'} = \sigma_{v'}$ for $v' \geq v$. In particular note that $\varphi_e(\sigma, \eta)_v = \sigma_v = y_v$. Then a simple consequence of the labeling gives that

$$P(\sigma)P(\eta) \leq P(u)P(\varphi_e(\sigma, \eta))e^{4\mathcal{E}(\beta)}$$

and a crude bound on the transition probabilities gives that

$$P(x \to y) \geq \frac{1}{n} \frac{e^{h_v y_v - d\beta}}{e^{h_v y_v - d\beta} + e^{-h_v y_v + d\beta}}.$$

Then

$$\sum_{\sigma,\eta:e\in\gamma(\sigma,\eta)} \frac{P(\sigma)P(\eta)}{Q(e)} \leq e^{4\mathcal{E}\beta} \frac{1}{P(x\to y)} \sum_{\sigma,\eta:e\in\gamma(\sigma,\eta)} P(\varphi_e(\sigma,\eta))$$
$$\leq ne^{4\mathcal{E}\beta}(1+e^{-2h_v y_v+2d\beta}) \sum_{\sigma,\eta:e\in\gamma(\sigma,\eta)} P(\varphi_e(\sigma,\eta)).$$

The labeling is constructed such that for each e the map φ_e is injective and as noted above we have that $\varphi_e(\sigma,\eta)_v = \sigma_v = y_v$ and so

$$\sum_{\sigma,\eta:e\in\gamma(\sigma,\eta)} P(\varphi_e(\sigma,\eta)) = \sum_{\sigma:\sigma_v=y_v} P(\sigma) \leq \frac{e^{hy_v+d\beta}}{e^{hy_v+d\beta}+e^{-hy_v-d\beta}} = \frac{1}{1+e^{-2hy_v-2d\beta}}$$

and hence

$$\tau \leq n^2 e^{4\mathcal{E}\beta} \frac{1+e^{-2h_v y_v+2d\beta}}{1+e^{-2h\sigma_v-2d\beta}} = n^2 e^{4\mathcal{E}\beta} \frac{1+e^{-2h_v\sigma_v}e^{2d\beta}}{1+e^{-2h\sigma_v}e^{-2d\beta}} \leq n^2 e^{4\mathcal{E}\beta+4d\beta}$$

as required. ■

We now need to establish a bound to relate the relaxation time to the mixing time. While we would like to apply equation (6) directly to Lemma 8, if the external fields go to infinity the right hand side of equation (6) also goes to infinity. So that our results holds for any external field we establish the following lemma.

Lemma 9 *Consider the Ising model on G with interaction strengths bounded by β, cut-width $\mathcal{E}$, arbitrary external field and maximal degree d. Then the mixing time of the Gibbs sampler satisfies,*

$$\tau_{mix} \leq 80n^3 e^{5\beta(\mathcal{E}+d)}.$$

Proof: Define $\bar{h} = 3\log n + 6\beta\mathcal{E} + 4d\beta + 10$ and let U denote the set of vertices $U = \{v \in V : |h_v| \geq \bar{h}\}$. These are the set of vertices with external fields so strong that it is highly unlikely that they are updated to a value other than sign(h_v). Let $\tilde{G}$ denote the graph induced by the vertex set $\tilde{V} = V \setminus U$, and let $\tilde{P}$ denote the Ising model with the same interaction strengths β_{uv} but with modified external field

$$\tilde{h}_v = h_v + \sum_{u\in U:(u,v)\in E} \beta_{uv}\,\text{sign}(h_u).$$

This is of course just the original Ising model restricted to $\tilde{V}$ with external field given by $\sigma_u = \text{sign}(h_u)$ for $u \in U$. We now analyze the continuous time Gibbs sampler of $\tilde{P}$. By Lemma 8 its relaxation time satisfies

$$\tilde{\tau} \leq ne^{4\beta(\mathcal{E}+d)}$$

since restricting to $\tilde{G}$ can only decrease the cut-width and maximum degree and since the discrete and continuous relaxation times differ by a factor of n. To invoke (6) we bound $\min_\sigma \tilde{P}(\sigma)$. By our construction we have that

$$\max_{v\in\tilde{V}} |\tilde{h}_v| \leq \bar{h} + d\beta.$$

Now

$$\min_{\sigma\in\{+,-\}^{\tilde{V}}} \tilde{H}(\sigma) = \min_\sigma \sum_{\{v,u\}\in\tilde{E}} \beta_{u,v}\sigma(v)\sigma(u) + \sum_{v\in\tilde{V}} h_v\sigma(v) \geq -n(2d\beta+\bar{h})$$

and similarly $\max_\sigma \tilde{H}(\sigma) \leq n(2d\beta+\bar{h})$. Now the normalizing constant $\tilde{Z}$ satisfies

$$\tilde{Z} = \sum_{\sigma\in\{+,-\}^{\tilde{V}}} \exp(\tilde{H}(\sigma)) \leq 2^n \exp(n(2d\beta+\bar{h}))$$

so finally

$$\min_{\sigma\in\{+,-\}^{\tilde V}} \tilde P(\sigma) \geq \frac{\min_\sigma \exp(\tilde H(\sigma))}{\tilde Z} \geq 2^{-n}\exp(-n(4d\beta+2\bar h)).$$

By equation (6) this implies that the mixing time of the continuous time Gibbs sampler on $\tilde P$ satisfies

$$\tilde\tau_{mix} \leq \tilde\tau\left(1+\frac{1}{2}\log(\min_\sigma \tilde P(\sigma)^{-1})\right) \leq ne^{4\beta(\mathcal{E}+d)}\left(1+\frac{1}{2}n(\log 2+2d\beta+\bar h)\right).$$

We set $T = 8n^2\bar h e^{4\beta(\mathcal{E}+d)} \geq 4\tilde\tau_{mix}$.

We now return to the continuous time dynamics on all G. Let $\mathcal{A}$ denote the event that every vertex in $u\in U$ is updated at least once before time T. The probability that a vertex u is updated by time T is $1-e^{-T}$ and so by a union bound

$$P(\mathcal{A}) \geq 1-ne^{-T} \geq 1-ne^{-\bar h} \geq 1-e^{-10}.$$

Let $\mathcal{B}$ be the event that for every vertex $u\in U$ every update up to time $2T$ updates the spin to sign(h_u). For a single vertex $u\in U$ and any configuration σ when u is updated,

$$P\left(u \text{ is updated to } -\text{sign}(h_u)\right) \leq \frac{e^{-|h_u|+d\beta}}{e^{-|h_u|+d\beta}+e^{|h_u|-d\beta}} \leq e^{-2\bar h+2d\beta} \tag{20}$$

The number of updates in U up to time $2T$ is distributed as a Poisson random variable with mean $2T|U|$ so

$$\begin{aligned}
P(\mathcal{B}) &\geq P(\text{Po}(2Tne^{-2\bar h+2d\beta})=0)\\
&= e^{-2Tne^{-2\bar h+2d\beta}}\\
&\geq 1-2Tne^{-2\bar h+2d\beta}\\
&\geq 1-8n^3\bar h e^{4\beta(\mathcal{E}+d)-2\bar h+2d\beta}\\
&= 1-8\bar h e^{-\bar h-10}\\
&> 1-8e^{-10}
\end{aligned}$$

where the last inequality follows from the fact that $e^x > x$.

Let X_t denote the Gibbs sampler with respect to P and let Y_t be its restriction to $\tilde V$. Conditioned on $\mathcal{A}$ and $\mathcal{B}$ by time T every vertex in U has been updated and it has been updated to sign(h_u) and remains with this spin until time $2T$. For $T\leq t\leq 2T$ let Y_t denote the Gibbs sampler on $\tilde V$ with respect to $\tilde P$ with initial condition $Y_T = X_T(\tilde V)$. From time T to $2T$ couple X_t and Y_t with the same updates (that is inside $\tilde V$ the same choice of $\{v_i\}$ and $\{U_i\}$in the notation of Section 2.1). Then conditioned on $\mathcal{A}$ and $\mathcal{B}$ we have that $Y_t = X_t(\tilde V)$ for $T\leq t\leq 2T$.

We can now use our bound on the mixing time of the Gibbs sampler with respect to $\tilde P$. Since $T\geq 4\tilde\tau_{mix}$ by equation (4) we have that,

$$\|P\left(Y_{2T}=\cdot\right)-\tilde P(\cdot)\|_{\text{TV}} \leq e^{-4}. \tag{21}$$

Under the stationary measure P it follows from equation (20) that for any $u\in U$,

$$P\left(\sigma_u = \text{sign}(h_u)\right) \geq 1-e^{2|h_u|-2d\beta}$$

and hence by a union bound

$$P\left(\sigma_u=\text{sign}(h_u),\ \forall u\in U\right) \geq 1-ne^{2\bar h-2d\beta}. \tag{22}$$

and so

$$\|P\left(\sigma\in\cdot \mid \sigma_u=\text{sign}(h_u),\ \forall u\in U\right)-P(\sigma\in\cdot)\|_{\text{TV}} \leq ne^{2\bar h-2d\beta}.$$

Since the projection of P onto $\tilde{V}$ conditioning on $\sigma_u = \text{sign}(h_u)$ for all $u \in U$ is simply $\tilde{P}$ it follows that

$$\begin{aligned}
\|P(X_{2T} = \cdot) - \tilde{P}(\cdot)\|_{\text{TV}} &\leq P(\mathcal{A}^c) + P(\mathcal{B}^c) + \|P(\sigma \in \cdot \mid \sigma_u = \text{sign}(h_u),\ \forall u \in U) - P(\sigma \in \cdot)\|_{\text{TV}} \\
&\quad + \|P(Y_{2T} \in \cdot) - \tilde{P}(\sigma \in \cdot)\|_{\text{TV}} \\
&\leq 9e^{-10} + ne^{2\bar{h} - 2d\beta} + e^{-4} \\
&\leq \frac{1}{2e}
\end{aligned}$$

which established $2T$ as an upper bound on the mixing time τ_{mix}. By a crude bound $\bar{h} \leq 10ne^{\beta(d+\mathcal{E})}$ which establishes

$$\tau_{mix} \leq 2T \leq 8n^2\bar{h}e^{4\beta(\mathcal{E}+d)} \leq 80n^3e^{5\beta(\mathcal{E}+d)}$$

as required.

■

4.2 Proof of Local Mixing for Graphs of Bounded Degree

We can now prove Lemma 2.

Proof: The proof follows immediately from Lemma 9 applied to the balls $B(v, R)$ noting that $\mathcal{E}$ is always smaller than the number of vertices in the graph which is bounded by $\mathfrak{X}$. ■

4.3 Cut-width in Random Graphs and Galton Watson Trees

The main result we prove in this section is the following.

Lemma 10 *For every d there exists a constant $C'(d)$ such that the following hold. Let T be the tree given by the first ℓ levels of a Galton-Watson branching process tree with Poisson(d) offspring distribution. Then $\mathcal{E}(T)$, the cut-width of T, is stochastically dominated by the distribution $C'\ell + \text{Po}(d)$.*

Using this result it is not hard to prove the upper bound on the local mixing of Lemma 5.

Proof: We first note that by Lemma 7 with high probability for all v, the tree excess of the ball $B(v, R)$ is at most one. This implies that the cut-width of $B(v, R)$ is at most 1 more than the cut-width of the spanning tree $T(v, R)$ of $B(v, R)$ whose distribution is dominates by a Galton-Watson tree with Poisson offspring distribution with mean d. We thus conclude by Lemma 10 that with high probability for all $v \in V$ the distribution of the cut-width of $B(v, R)$ is bounded by $C'R + \text{Po}(d)$. Since the probability that $\text{Po}(d)$ exceeds $c \log n / \log\log n$ for large enough c is of order n^{-2}, we obtain by a union bound that with high probability for all v it holds that $B(v, R)$ has a cut-width of at most $(c + C') \log n / \log\log n$. Similarly with high probability the maximal degree in G is of order $\log n / \log\log n$. Recalling that $\mathfrak{X}$ is at most $d^R \log n$ and applying Lemma 9 yields the required result. ■

The proof of Lemma 10 follows by induction from the following two lemmas.

Lemma 11 *Let T be a tree rooted at ρ with degree m and let $T_1, \ldots, T_k$ be the subtrees connected to the root. Then the cut-width of T satisfies,*

$$\mathcal{E}(T) \leq \max_i \mathcal{E}(T_i) + k + 1 - i.$$

Proof: For each subgraph T_i let $u_1^{(i)}, \ldots, u_{|V_i|}^{(i)}$ be a sequence on vertices which achieves the cut-width $\mathcal{E}(T_i)$. Concatenate these sequences as

$$\rho, u_1^{(1)}, \ldots, u_{|V_1|}^{(1)}, u_1^{(2)}, \ldots, u_{|V_k|}^{(k)}$$

which can easily be seen to achieve the bound $\max_i \mathcal{E}(T_i) + k + 1 - i$. ■

For a collection of random variables $Y_1, \ldots, Y_k$ the order statistics is defined as the permutation of the values into increasing order such that $Y_{(1)} \leq \ldots \leq Y_{(k)}$.

Lemma 12 *Let $X \sim$ Po(d) and let $Y_1, \ldots, Y_X$ be an iid sequence distributed as Po(d). There exists $C(d)$ such that*

$$W = X + \max_{1 \leq i \leq X} Y_{(i)} - i$$

is stochastically dominated by $C +$ Po(d).

Proof: The probability distribution of the Poisson is given by $P(\text{Po}(d) = w) = \frac{d^w e^{-d}}{w!}$ which decays faster than any exponential so

$$\frac{P(\text{Po}(d) \geq w)}{P(\text{Po}(d) = w)} \to 1$$

as $w \to \infty$. With this fast rate of decay we can choose $C = C(d)$ large enough so that the following hold:

- That $C \geq 6$ is even and for $w \geq \frac{C}{2}$,
$$P(\text{Po}(d) \geq w + 1) \leq P(\text{Po}(d) = w) \tag{23}$$
- For all $w \geq 0$,
$$\left(w + \frac{C}{2}\right) E2^X P(\text{Po}(d) \geq w + \frac{C}{2}) \leq \frac{1}{100} P(\text{Po}(d) \geq w) \tag{24}$$
- For all $w \geq 0$,
$$P(\text{Po}(d) \geq \lfloor \frac{w}{2} \rfloor + C)^3 \leq P(\text{Po}(d) \geq w + \frac{C}{2}) \tag{25}$$
which can be achieved since $\frac{1}{((\lfloor \frac{w}{2} \rfloor + C)!)^3} << \frac{1}{(w + \frac{C}{2})!}$.
- For all $w \geq 2$,
$$\left(w + \frac{C}{2}\right)^2 2^{2w + \frac{3C}{2}} P\left(\text{Po}(d) \geq \frac{C}{2}\right)^{\lfloor \frac{w}{2} \rfloor} \leq \frac{1}{100} \tag{26}$$
- For $w \in \{0, 1\}$,
$$P(W \geq w + C) \leq P\left(\text{Po}(d) \geq w\right) \tag{27}$$

Observe that for $1 \leq i \leq x$,

$$P(Y_{(i)} \geq w \mid X = x) \leq \binom{x}{x - i + 1} P(\text{Po}(d) \geq w)^{x-i+1} \leq 2^x P(\text{Po}(d) \geq w)^{x-i+1} \tag{28}$$

since if $Y_{(i)} \geq w$ then there are at least $x - i + 1$ of the Y's must be greater than or equal to w and there are $\binom{x}{x-i+1}$ such choices of the set. For any $y, z \geq 0$ we have that

$$P(\text{Po}(d) = y) P(\text{Po}(d) = z) = \frac{d^y e^{-d}}{y!} \frac{d^z e^{-d}}{z!} = \binom{y+z}{z} \frac{d^{y+z} e^{-2d}}{(y+z)!} \leq 2^{y+z} P(\text{Po}(d) = y + z) \tag{29}$$

since $\binom{y+z}{z} \leq 2^{y+z}$.

Fix a $w \geq 2$. Then

$$\begin{aligned}
P(W \geq w+C) &= P(X + \max_{1\leq i\leq X} Y_{(i)} - i \geq w+C) \\
&\leq P(X > w + \frac{C}{2}) + \sum_{x=1}^{w+\frac{C}{2}} P(x + \max_{1\leq i\leq x} Y_{(i)} - i \geq w + C \mid X = x)P(X=x) \\
&\leq \frac{1}{100}P(X=w) + \sum_{x=1}^{w+\frac{C}{2}} P(x + \max_{1\leq i\leq x} Y_{(i)} - i \geq w + C \mid X = x)P(X=x) \qquad (30)
\end{aligned}$$

where the final equality follows from equation (24). Now

$$\begin{aligned}
&\sum_{x=1}^{w+\frac{C}{2}} P(x + \max_{1\leq i\leq x} Y_{(i)} - i \geq w + C \mid X = x)P(X=x) \\
&\quad\leq \sum_{x=1}^{w+\frac{C}{2}} \sum_{i=1}^{x} P(x + Y_{(i)} - i \geq w + C \mid X = x)P(X=x) \\
&\quad= \sum_{x=1}^{w+\frac{C}{2}} \sum_{j=1}^{x} P(Y_{(x-j+1)} \geq w - j + 1 + C \mid X = x)P(X=x) \\
&\quad\leq \sum_{x=1}^{w+\frac{C}{2}} \sum_{j=1}^{x} 2^x P(\text{Po}(d) \geq w - j + 1 + C)^j P(X=x) \\
&\quad= \sum_{j=1}^{w+\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^x P(\text{Po}(d) \geq w - j + 1 + C)^j P(X=x) \qquad (31)
\end{aligned}$$

where line 3 follows by setting $j = x - i + 1$ and line 4 follows from equation (28). We split this sum into 3 parts. First we have that

$$\begin{aligned}
&\sum_{j=1}^{\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^x P(\text{Po}(d) \geq w - j + 1 + C)^j P(X=x) \\
&\quad\leq \frac{C}{2} \sum_{x=1}^{w+\frac{C}{2}} 2^x P(\text{Po}(d) \geq w + \frac{C}{2})P(X=x) \\
&\quad\leq \frac{C}{2} E2^X P(\text{Po}(d) \geq w + \frac{C}{2}) \\
&\quad\leq \frac{1}{100} P(\text{Po}(d) \geq w) \qquad (32)
\end{aligned}$$

where the final equality follows from equation (24). Second,

$$\begin{aligned}
&\sum_{j=\frac{C}{2}+1}^{\lfloor \frac{w}{2} \rfloor} \sum_{x=j}^{w+\frac{C}{2}} 2^x P(\text{Po}(d) \geq w-j+1+C)^j P(X=x) \\
&\quad \leq \lfloor \frac{w}{2} \rfloor \sum_{x=\frac{C}{2}+1}^{w+\frac{C}{2}} 2^x P(\text{Po}(d) \geq \lfloor \frac{w}{2} \rfloor + C)^{\frac{C}{2}} P(X=x) \\
&\quad \leq \lfloor \frac{w}{2} \rfloor E 2^X P(\text{Po}(d) \geq \lfloor \frac{w}{2} \rfloor + C)^{\frac{C}{2}} \\
&\quad \leq \lfloor \frac{w}{2} \rfloor E 2^X P(\text{Po}(d) \geq w + \frac{C}{2}) \\
&\quad \leq \frac{1}{100} P(\text{Po}(d) \geq w)
\end{aligned} \tag{33}$$

where line 4 follows from the fact that $\frac{C}{2} \geq 3$ and equation (25) and line 5 follows from equation (24). Finally,

$$\begin{aligned}
&\sum_{j=\lfloor \frac{w}{2} \rfloor+1}^{w+\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^x P(\text{Po}(d) \geq w-j+1+C)^j P(X=x) \\
&\quad \leq \sum_{j=\lfloor \frac{w}{2} \rfloor+1}^{w+\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^{w+\frac{C}{2}} P(\text{Po}(d) \geq w-j+1+C)^{\lfloor \frac{w}{2} \rfloor+1} P(\text{Po}(d)=x) \\
&\quad \leq \sum_{j=\lfloor \frac{w}{2} \rfloor+1}^{w+\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^{w+\frac{C}{2}} P(\text{Po}(d) \geq \frac{C}{2})^{\lfloor \frac{w}{2} \rfloor} P(\text{Po}(d)=w-x+C) P(\text{Po}(d)=x)
\end{aligned} \tag{34}$$

where the second line follows since $x \leq w + \frac{C}{2}$ and $j \geq \lfloor \frac{w}{2} \rfloor + 1$ and the third line follows from the fact that $w - j + 1 + C$ is greater than both $\frac{C}{2}$ and $w - x + C + 1$ and applying equation (23) which says that $P(\text{Po}(d) = w - x + C) \geq P(\text{Po}(d) \geq w - x + C + 1)$. Then

$$\begin{aligned}
&\sum_{j=\lfloor \frac{w}{2} \rfloor+1}^{w+\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^{w+\frac{C}{2}} P(\text{Po}(d) \geq \frac{C}{2})^{\lfloor \frac{w}{2} \rfloor} P(\text{Po}(d)=w-x+C) P(\text{Po}(d)=x) \\
&\quad \leq \sum_{j=\lfloor \frac{w}{2} \rfloor+1}^{w+\frac{C}{2}} \sum_{x=j}^{w+\frac{C}{2}} 2^{w+\frac{C}{2}} P(\text{Po}(d) \geq \frac{C}{2})^{\lfloor \frac{w}{2} \rfloor} 2^{w+C} P(\text{Po}(d)=w+C) \\
&\quad \leq \left(w + \frac{C}{2}\right)^2 2^{2w+\frac{3C}{2}} P(\text{Po}(d) \geq \frac{C}{2})^{\lfloor \frac{w}{2} \rfloor} P(\text{Po}(d)=w+C) \\
&\quad \leq \frac{1}{100} P(\text{Po}(d) \geq w)
\end{aligned} \tag{35}$$

where the second line follows from equation (29), and the final line follows from equation (26). Combining equations (30) through (35) we have that for $w \geq 2$,

$$P(W \geq w + C) \leq \frac{1}{25} P\left(\text{Po}(d) \geq w\right) \leq P\left(\text{Po}(d) \geq w\right).$$

Combining this with equation (27) completes the proof. ■

We now prove Lemma 10.

Proof: Take $C' = C + 1$ where C is the constant from Lemma 12. We prove the result by induction on ℓ. When $\ell = 0$ a 0 level Galton-Watson branching process tree is just a single vertex which has cut-width 0 so the statement is trivially satisfied. When $\ell \geq 1$ the subtrees attached to the root are independent $\ell - 1$ level Galton-Watson branching process trees so by the inductive hypothesis, Lemma 11 and Lemma 12 we have that $\mathcal{E}(T)$ is stochastically dominated by the distribution $C'\ell + \mathrm{Po}(d)$. ■

5 Spatial Mixing

5.1 SAW trees

Weitz [30] developed the Tree of Self Avoiding Walks construction which enables the calculation of marginal distributions of a Gibbs measure on a graph by calculating marginal distributions on a specially constructed tree. This construction, along with the censoring inequality, will be a major tools in our proof. For a graph G and a vertex v we denote the tree of self-avoiding paths from V in G as $T_{saw}(G, v)$. This is the tree of paths in G starting from v and not intersecting themselves, except possibly at the terminal vertex of the path. Through this construction each vertex in $T_{saw}(G, v)$ can be identified with a vertex in G which gives a natural way to relate a subset $\Lambda \subset V$ and a configuration η_Λ to the corresponding subset and configuration in $T_{saw}(G, v)$ which we denote $\varphi(\Lambda) \subset T_{saw}(G, v)$ and $\eta_{\varphi(\Lambda)}$ respectively. Furthermore if $A, B \subset V$ then $d(A, B) = d(\varphi(A), \varphi(B))$. Each vertex (edge) of T_{saw} corresponds to a vertex (edge) in $T_{saw}(G, v)$ so $P_{T_{saw}}$ is defined by taking the corresponding external field and interactions. Then Theorem 3.1 of [30] gives the following result.

Lemma 13 *[Weitz [30]] For a graph G and $v \in G$ there exists $A \subset T_{saw}$ and a configuration ν_A on A such that for any $\lambda \subset V$ and configuration η_Λ on Λ such that,*

$$P_G(\sigma_v = +|\sigma_\Lambda) = P_{T_{saw}}(\sigma_v = +|\sigma_{\varphi(\Lambda)\setminus A} = \eta_{\varphi(\Lambda)\setminus A}, \sigma_A = \nu_A).$$

The set A is the set of leaves in T_{saw} corresponding to the terminal vertices of paths which return to a vertex already visited by the path. The construction of ν_A is described in [30].

5.2 Spatial Correlations on trees

We consider the effect that conditioning the vertices of a tree has on the marginal distribution of the spin at the root. It will be convenient to compare this probability to the Ising model with the same interaction strengths β_{uv} but no external field ($h \equiv 0$) which we will denote $\widetilde{P}$.

Lemma 14 *Suppose that T is a tree, P is the Ising model with arbitrary external field (including $h_u = \pm\infty$ meaning that σ_u is set to $\pm$) and $0 \leq \beta_{u,v} \leq \beta$ for all $(u, v) \in E$. Let $U \subseteq \Lambda \subset V$, and let η^+, η^- be two configurations on Λ which differ only on U with $\eta^+_U \equiv +, \eta^-_U \equiv -$. Then for all $v \in V$,*

$$0 \leq P(\sigma_v = +|\sigma_\Lambda = \eta^+) - P(\sigma_v = +|\sigma_\Lambda = \eta^-) \leq \sum_{u\in U} (\tanh \beta)^{d(u,v)}.$$

Proof: The inequality

$$0 \leq P(\sigma_v = +|\sigma_\Lambda = \eta^+) - P(\sigma_v = +|\sigma_\Lambda = \eta^-)$$

simply follows from the monotonicity of the ferromagnetic Ising model. Now suppose that the set U is a single vertex u. Lemma 4.1 of [2] implies that for any vertices $v, u \in T$,

$$P(\sigma_v = +|\sigma_u = +) - P(\sigma_v = +|\sigma_u = -) \leq \widetilde{P}(\sigma_v = +|\sigma_u = +) - \widetilde{P}(\sigma_v = +|\sigma_u = -). \tag{36}$$

If $u_0, u_1, \ldots, u_l$ are a path of vertices in T then a simple calculation yields that

$$\widetilde{P}(\sigma_{u_k} = +|\sigma_{u_0} = +) - \widetilde{P}(\sigma_{u_k} = +|\sigma_{u_0} = -) = \prod_{i=1}^{k} \tanh \beta_{u_{i-1}u_i} \leq (\tanh \beta)^k. \tag{37}$$

Conditioning is equivalent to setting an infinite external field so equations (36) and (37) imply that

$$P(\sigma_v = +|\sigma_\Lambda = \eta^+) - P(\sigma_v = +|\sigma_\Lambda = \eta^-) \leq (\tanh \beta)^{d(u,v)}. \tag{38}$$

We now consider a general U. Let $u_1, \ldots, u_{|U|}$ be an arbitrary labeling of the vertices of U. Take a sequence of configurations $\eta^0, \eta^1, \ldots, \eta^{|U|}$ on Λ with $\eta^0 = \eta^-$ and $\eta^{|U|} = \eta^+$ where consecutive configurations η^{i-1} and η^i differ only at u_i with $\eta^i_{u_i} = +$ and $\eta^{i-1}_{u_i} = -$. By equation (38) we have that

$$P(\sigma_v = +|\sigma_\Lambda = \eta^{i+1}) - P(\sigma_v = +|\sigma_\lambda = \eta^i) \leq (\tanh \beta)^{d(v,u_i)}$$

and so

$$P(\sigma_v = +|\sigma_\Lambda = \eta^+) - P(\sigma_v = +|\sigma_\Lambda = \eta^-) \leq \sum_{u \in U} (\tanh \beta)^{d(u,v)}$$

which completes the proof.

■

5.3 Continuous time to discrete time

Lemma 15 *Suppose that in continuous time starting from the all $+$ and all $-$ configurations the Gibbs sampler under the monotone coupling couples with probability at least $\frac{7}{8}$ by time $T \geq 1$. Then the Gibbs sampler in discrete time under the monotone coupling couples with probability at least $1 - \frac{1}{2e}$ by time $\lceil 5Tn \rceil$ and hence has mixing time at most $\lceil 5Tn \rceil$.*

Proof: Let M denote the number of updates of the continuous dynamics up to time T. Then M is distributed as a Poisson random variable with mean Tn. For some integer m, the final state of the continuous time Gibbs sampler conditioned on $M = m$ is the same as the final state of the discrete Gibbs sampler with m steps. So the probability of coupling in the discrete time after m steps is at least $\frac{7}{8} - P(\text{Po}(Tn) > m)$. So if $m \geq 5Tn$ then by Markov's Theorem

$$P(\text{Po}(Tn) > m) \leq \frac{E e^{\text{Po}(Tn)}}{e^{5Tn}} = e^{Tn(e-1) - 5Tn} \leq e^{-3}.$$

Since $\frac{7}{8} - e^{-3} > 1 - \frac{1}{2e}$ the discrete chain couples by time $5Tn$ with probability at least $1 - \frac{1}{2e}$. Hence the mixing time is at most $\lceil 5Tn \rceil$. ■

5.4 Proof of Lemma 3

We now prove Lemma 3 by applying Lemma 13 and 14 to a small graph centered at v.

Proof:

Let T denote the tree of self avoiding walks on G from v, $T_{saw}(G, v)$. Let $\varphi(S(v, R))$ denote the vertices in T which correspond to vertices in $S(v, R)$ and for each $u \in S(v, R)$ let $\varphi(u)$ denote the set of vertices in T

which correspond to u. Then by Lemma 13 and Lemma 14,

$$\begin{aligned} a_u &= \sup_{\eta^+,\eta^-} P_{T_{saw}}(\sigma_v = +|\sigma_{\varphi(\Lambda)\setminus A} = \eta^+_{\phi(\Lambda)\setminus A}, \sigma_A = \nu_A) - P_{T_{saw}}(\sigma_v = +|\sigma_{\varphi(\Lambda)\setminus A} = \eta^-_{\phi(\Lambda)\setminus A}, \sigma_A = \nu_A) \\ &\leq \sum_{w\in\varphi(u)} \tanh^{d(v,w)} \beta. \end{aligned} \tag{39}$$

Applying this bound

$$\begin{aligned} \sum_{u\in S(v,R)} a_u &\leq \sum_{u\in S(v,R)} \sum_{w\in\varphi(u)} \tanh^{d(v,w)} \beta \\ &= \sum_{w\in\varphi(S(v,R))} \tanh^{d(v,w)} \beta \\ &\leq \sum_{w\in T:d(w,v)\geq R} \tanh^{d(v,w)} \beta \end{aligned}$$

where the final inequality follows from the fact that $d(v,\varphi(S(v,R))) \geq m$. Now since T has maximum degree d for each ℓ there are at most $d(d-1)^{\ell-1}$ vertices at distance ℓ from v. It follows that

$$\begin{aligned} \sum_{u\in S(v,R)} a_u &\leq \sum_{w\in T:d(w,v)\geq R} \tanh^{d(v,w)} \beta \\ &\leq \sum_{\ell=R}^{\infty} d(d-1)^{\ell-1} \tanh^{\ell} \beta \\ &= \frac{d(d-1)^{R-1}\tanh^R \beta}{1-(d-1)\tanh\beta} \\ &\leq \frac{1}{4} \end{aligned}$$

as required. ■

5.5 Proof of Lemma 6

We now prove Lemma 6.

Proof:

We need to establish the spatial mixing condition. Recall that

$$a_u = \sup_{\eta^+,\eta^-} P(\sigma_v = +|\sigma_\Lambda = \eta^+) - P(\sigma_v = +|\sigma_\Lambda = \eta^-)$$

and by equation (39)

$$a_u \leq \sum_{w\in\varphi(u)} \tanh^{d(v,w)} \beta.$$

Now $t(v,R) \leq 1$ with high probability for all $v \in V$ by Lemma 7 so $B(v,R)$ is a tree or unicyclic. Hence every $u \in S(v,R)$ appears at most twice in the tree of self-avoiding walks which gives $|\varphi(u)| \leq 2$ and

$d(v,\varphi(u)) = R$. Thus for all $v \in V$ with high probability

$$\begin{aligned}\sum_{u\in S(v,R)} a_u &\le \sum_{u\in S(v,R)} \sum_{w\in\varphi(u)} \tanh^{d(v,w)}\beta \\ &\le 2\mathfrak{X}\tanh^R\beta \\ &= 6(1-d^{-1})^{-1}(d\tanh\beta)^R\log n \\ &= o(1)\end{aligned}$$

which establishes the spatial mixing condition.

■

6 Infinite Graphs

Up until this point we have only dealt with finite graphs, however, the Ising model and the Glauber dynamics can be defined on infinite graphs as well, (see e.g. [14]). The spatial mixing property of uniqueness says that there is a unique Gibbs measure for the interacting particle system; one formulation of this is that for every finite set $A \subset V$ we have that

$$\limsup_{R\to\infty} \sup_{\eta,\eta'} \| P\left(\sigma_A = \cdot \mid \sigma_{S(A,R)} = \eta\right) - P\left(\sigma_A = \cdot \mid \sigma_{S(A,R)} = \eta'\right) \|_{\mathrm{TV}} = 0$$

where $S(A,R) = \{u \in V : d(u,A) = R\}$ and η, η' are configurations on $S(A,R)$. This says that the configuration on A is asymptotically independent of the spins a large distance away. In the context of the ferromagnetic Ising model this is equivalent to,

$$P\left(\sigma_v = + \mid \sigma_{S(v,R)} \equiv +\right) - P\left(\sigma_v = + \mid \sigma_{S(v,R)} \equiv -\right) \longrightarrow 0 \tag{40}$$

for all $v \in V$ as $R \to \infty$. Combining Lemmas 13 and 14 it follows that Condition (1) implies uniqueness. This was also noted in [**?**].

The following lemma shows that given uniqueness the Glauber dynamics on an infinite graph can locally be approximated by the Glauber dynamics of the Ising model on finite graphs. For a fixed finite set $U \subset$ let $\sigma^{*\ell}$ denote a random configuration according to the stationary distribution of the Ising model on the induced subgraph G_ℓ whose vertex set is given by $U_\ell := \{u \in V : d(u,U) \le \ell\}$. Let $\sigma^{*\ell}(t)$ denote the Glauber dynamics of this Ising model started from the stationary distribution.

Lemma 16 *Let G be a infinite graph with maximum degree d and suppose for some $\{\beta_{(u,v)}\}$ and $\{h_u\}$ the Ising model has the uniqueness property and let U be a finite subset of V. With $\sigma_U^{*\ell}(t)$ defined as above*

$$\left(\sigma_U^{*\ell}(0), \sigma_U^{*\ell}(1)\right) \to (\sigma_U(0), \sigma_U(1))$$

jointly in distribution as $\ell \to \infty$.

Proof: Fix an $\epsilon > 0$. It is sufficient to show that for some ℓ' we can couple $\left(\sigma_U^{*\ell}(0), \sigma_U^{*\ell}(1)\right)$ and $(\sigma_U(0), \sigma_U(1))$ with probability at least $1-\epsilon$ when $\ell > \ell'$. Fix some positive integer m large enough so that

$$P(\text{Poison}(1) \ge m) < \frac{1}{2}\epsilon d^{-m}|U|^{-1}.$$

By the uniqueness property as $\ell \to \infty$ we have that $\sigma_{U_m}^{*\ell}$ converges in distribution to σ_{U_m}. So for some ℓ' when $\ell > \ell'$ we can couple initial configurations $\sigma^{*\ell}(0)$ and $\sigma(0)$ so that $\sigma_{U_m}^{*\ell}(0)$ and $\sigma_{U_m}(0)$ agree with probability

at least $1-\epsilon/2$. Now couple the Glauber dynamics by using the same sequence of updates for each chain within U_ℓ .

We now bound the probability that there is disagreement between $\sigma_U^{*\ell}(1)$ and $\sigma_U(1)$ given that $\sigma_{U_m}^{*\ell}(0)$ and $\sigma_{U_m}(0)$ agree. We will call a sequence $u_1,\ldots,u_k$ of vertices a *path* if u_i and u_{i+1} are adjacent for each i. An update can only create a disagreement at the vertex if a neighboring vertex already has a disagreement. Hence a vertex u can only have a disagreement by time t if there is a path of vertices from $u_1,\ldots,u_k=u$ such that the vertices in the path are updated by the Glauber dynamics in that order before time 1 and $u_1\in U_m\setminus U_{m-1}$.

Hence the event $\sigma_U^{*\ell}(1)\neq\sigma_U(1)$ is dominated by the event that there is a path of updates of vertices $u_1,\ldots,u_m$, updated in that order before time 1 with $u_m\in U$. For each fixed path the probability that those vertices are updated in that order is $P(\text{Poison}(1)\geq m)$. There are at most $d^m|U|$ such paths of vertices so by a union bound and our choice of m the probability of a disagreement reaching $|U|$ is at most $\epsilon/2$. It follows that we can couple $\left(\sigma_U^{*\ell}(0),\sigma_U^{*\ell}(1)\right)$ and $(\sigma_U(0),\sigma_U(1))$ with probability at least $1-\epsilon$ which completes the proof. ■

We now show how the spectral gap bounds for the finite graph dynamics imply spectral gap bounds for infinite graph dynamics. The following lemma completes Theorem 1.

Lemma 17 *Let G be a infinite graph with maximum degree d and suppose for some $\{\beta_{(u,v)}\}$ and $\{h_u\}$ the Ising model has the uniqueness property. Further suppose that for every finite subgraph G' of G the Ising model on G' has continuous time spectral gap bounded below by λ^*. Then the infinite volume dynamics has spectral gap bounded below by λ^*.*

Proof: First we may assume that the graph is connected since the spectral gap is the minimum of the spectral gaps of the dynamics projected onto individual components. We will use the characterization of the spectral gap that

$$\text{Gap}=-\log\sup_f\frac{\text{Cov}((f(\sigma(0)),f(\sigma(1)))}{\text{Var}f(\sigma(0))}$$

where the supremum is over all square integrable functions $f:\{+,-\}^V\to\mathbb{R}$ with $Ef=0$. Fix a vertex v and for such a function f we define the bounded function $f_R:\{+,-\}^{B(v,R)}\to\mathbb{R}$ by

$$f_R(\sigma)=E\left(f(\sigma)\mid\sigma_{B(v,R)}\right)$$

Since every vertex is ultimately in $B(v,R)$ for R sufficiently large, by the L^2 Martingale Convergence Theorem $f_R(\sigma)$ converges to $f(\sigma)$ in L^2 and so

$$\lim_{R\to\infty}\frac{\text{Cov}((f_R(\sigma(0)),f_R(\sigma(1)))}{\text{Var}f_R(\sigma(0))}=\frac{\text{Cov}((f(\sigma(0)),f(\sigma(1)))}{\text{Var}f(\sigma(0))}.\tag{41}$$

In particular this means that in the supremum we only need consider bounded functions which are determined by a finite number of spins. So suppose that g is such a bounded function depending only on σ_U for some finite $U\subset V$.

By Lemma 16 we have that $\left(\sigma_U^{*\ell}(0),\sigma_U^{*\ell}(1)\right)$ converges jointly in distribution to $(\sigma_U(0),\sigma_U(1))$. Hence using our assumption on the spectral gap on finite subgraphs we have that

$$\lambda^*\leq\lim_{\ell\to\infty}-\log\frac{\text{Cov}\left(\left(g(\sigma^{*\ell}(0)),g(\sigma^{*\ell}(1)\right)\right)}{\text{Var}g(\sigma^{*\ell}(0))}=-\log\frac{\text{Cov}((g(\sigma(0)),g(\sigma(1)))}{\text{Var}g(\sigma(0))},$$

which establishes λ^* as a lower bound on the spectral gap. ■

7 Conclusion

The proof of Theorem 3 naturally extends to more general monotone systems. Moreover instead of censoring outside a ball of radius R about a vertex v we could instead look at general well chosen sets $v \in W_v \subset V$. We let S_v denote the boundary set $\{u \in V \setminus W_v : d(u, W_v) = 1\}$. We consider the following setup. There is a spin set Ω which is ordered with a maximal element $+$ and a minimal element $-$. The order on Ω naturally extends to a partial order on Ω^V where V is the vertex set of a graph by letting $\sigma_1 \leq \sigma_2$ if and only if $\sigma_1(v) \leq \sigma_2(v)$ for all $v \in V$. A measure P on Ω^V is called monotone if for all $v \in V$ and all $a \in \Omega$

$$P[\sigma(v) \geq a | \sigma(w : w \neq v) = \sigma_1] \geq P[\sigma(v) \geq a | \sigma(w : w \neq v) = \sigma_2],$$

whenever $\sigma_1 \geq \sigma_2$. We may now state a generalization of Theorem 3.

Theorem 4 *Let G be a graph on $n \geq 2$ vertices and let $P(\sigma)$ be any monotone Gibbs measure on G.*

Suppose that there exist constants $T, \mathfrak{X} \geq 1$ and for each $v \in V$ there is a subset $W_v \subset V$ containing v such that the following three conditions hold::

- ***Volume:*** *The volume of W_v satisfies $|W_v| \leq \mathfrak{X}$.*
- ***Local Mixing:*** *For any configuration η on S_v the continuous time mixing time of the Gibbs sampler on W_v with fixed boundary condition η is bounded above by T.*
- ***Spatial Mixing:*** *For each vertex $u \in S_v$ define*

$$a_u = \sup_{\eta^+, \eta^-} d_{TV} \left(P(\sigma_v = \cdot | \sigma_\Lambda = \eta^1), P(\sigma_v = \cdot | \sigma_\Lambda = \eta^2) \right) \tag{42}$$

where the supremum is over configurations η^1, η^2 on S_v which differ only at u. Then

$$\sum_{u \in S_v} a_u \leq \frac{1}{4}. \tag{43}$$

Then starting from the all $+$ and all $-$ configurations in continuous time the monotone coupling couples with probability at least $\frac{7}{8}$ by time $T \lceil \log 8\mathfrak{X} \rceil (3 + \log_2 n)$.

It follows that the mixing time of the Gibbs sampler in continuous time satisfies

$$\tau_{mix} \leq T \lceil \log 8\mathfrak{X} \rceil (3 + \log_2 n).$$

While Theorem 4 applies to general monotone systems, the use of the Censoring Lemma of Peres and Winkler does not allow to extend it to non-monotone systems such as random colourings. A major open problem is to relate spatial mixing to temporal mixing in non-monotone settings, for example for the hardcore model, antiferromagnetic Ising model, or colouring model.

7.1 Open Problems

We showed that Condition (1) establishes a uniform lower bound on the spectral gap of the continuous time dynamics over all graphs. It would be of interest to establish whether or not this is also true for bounds on the Log-Sobolev constant as well.

As discussed in the introduction our results give rise to the following conjecture concerning non-monotone systems.

Conjecture 1 *The Gibbs sampler for the antiferromagnetic Ising model (with no external field) is rapidly mixing on any graph whose maximum degree d, for any inverse temperature β below the uniqueness threshold for the Ising model on the d-regular tree.*

Similarly the Gibbs sampler for the hardcore model is rapidly mixing on any graph whose maximum degree is d for any fugacity λ below the uniqueness threshold for the hard-core model on the d-regular tree.

We recall that for both of these models the mixing time on almost all random d-regular bipartite graphs is exponential in n the size of the graph beyond the uniqueness threshold [25, 8, 5] so our conjecture is that uniqueness on the tree exactly corresponds to rapid mixing of the Gibbs sampler. A similar conjecture can be made with respect to the coloring model.

7.2 Acknowledgments

Most of this work was carried out while AS was a student at UC Berkeley and on visits to Rome Tre and the Weizmann Institute. AS would like to thank Fabio Martinelli for his useful discussions.

References

[1] D. Aldous and J. A. Fill. Reversible Markov chains and random walks on graphs. book in preparation. Current version online at http://stat-www.berkeley.edu/users/aldous/book.html.

[2] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Related Fields*, 131(3):311–340, 2005.

[3] F. Cesi. Quasi-factorization of the entropy and logarithmic sobolev inequalities for gibbs random elds, 2001.

[4] M. F. Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In *Probability towards 2000*, number 128 in Lecture Notes in Statist. 1998.

[5] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. submitted.

[6] R. L. Dobrushin and S. B. Shlosman. Constructive criterion for uniqueness of a Gibbs field. In J. Fritz, A. Jaffe, and D. Szasz, editors, *Statistical Mechanics and dynamical systems*, volume 10, pages 347–370. 1985.

[7] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: a combinatorial view. *Random Structures Algorithms*, 24(4):461–479, 2004.

[8] A. Gershchenfeld and A. Montanari. Reconstruction for models on random graphs. In *Foundations of Computer Science, Annual IEEE Symposium on*, pages 194–204. IEEE Computer Society, 2007.

[9] T. Hayes and E. Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, 2005.

[10] T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005)*, pages 511–520, 2005.

[11] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[12] C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 568–578. IEEE Computer Soc., Los Alamitos, CA, 2001.

[13] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

[14] Thomas M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[15] R. Lyons. The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys.*, 125(2):337–353, 1989.

[16] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

[17] F. Martinelli, Alistair A. Sinclair, and D. Weitz. Glauber dynamics on trees: boundary conditions and mixing time. *Comm. Math. Phys.*, 250(2):301–334, 2004.

[18] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.*, 161(3):447–486, 1994.

[19] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Comm. Math. Phys.*, 161(3):487–514, 1994.

[20] F. Martinelli, A. Sinclair, and D. Weitz. The ising model on trees: Boundary conditions and mixing time. In *Proceedings of the Forty Fourth Annual Symposium on Foundations of Computer Science*, pages 628–639, 2003.

[21] M. Mézard and A. Montanari. *Information, physics, and computation*. Oxford University Press, USA, 2009.

[22] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Clusters of solutions and replica symmetry breaking in random k-satisfiability. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P04004, 2008.

[23] E. Mossel and A. Sly. Rapid mixing of gibbs sampling on graphs that are sparse on average. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 238–247, 2008.

[24] E. Mossel and A. Sly. Rapid mixing of gibbs sampling on graphs that are sparse on average. To Appear In Random Structures and Algorithms, 2009.

[25] E. Mossel, D. Weitz, and N. Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Prob. Theory Related. Fields*, 143:401–439, 2009.

[26] Y. Peres. Mixing for Markov chains and spin systems. Draft lecture notes available at http://www.stat.berkeley.edu/

[27] D. W. Stroock and B. Zegarliński. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.*, 149(1):175–193, 1992.

[28] E. Vigoda. Improved bounds for sampling coloring. In *40th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 51–59, 1999.

[29] E. Vigoda. Improved bounds for sampling coloring. *Journal of Mathematical Physics*, (3):1555–1569, 2000.

[30] D. Weitz. Counting indpendent sets up to the tree threshold. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 140–149. ACM, 2006.

[31] Dror Weitz. Combinatorial criteria for uniqueness of gibbs measures. *Random Struct. Algorithms*, 27:445–475, 2005.